

## Durham Research Online

---

### Deposited in DRO:

08 September 2020

### Version of attached file:

Accepted Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Li, Shisong and Wang, Qing and Zhao, Wei and Huang, Songling (2020) 'From 0 to e : a survey of major impacts for electrical measurements in recent SI revision.', IEEE transactions on instrumentation and measurement., 69 (9). pp. 5956-5965.

### Further information on publisher's website:

<https://doi.org/10.1109/TIM.2020.3010351>

### Publisher's copyright statement:

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

### Additional information:

## Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

# The Relationship Between Curvilinear Structure Enhancement And Ridge Detection Methods

Haifa F. Alhasson · Chris G. Willcocks · Shuaa S. Alharbi · Adetayo Kasim · Boguslaw Obara

Received: date / Accepted: date

**Abstract** Curvilinear structure detection and quantification is a large research area with many imaging applications in fields such as biology, medicine, and engineering. Curvilinear enhancement is often used as a pre-processing stage for ridge detection, but there has been little investigation into the relationship between enhancement and ridge detection. In this paper, we thoroughly evaluate the pair-wise combinations of different curvilinear enhancement and ridge detection methods across two highly varied datasets, as well as samples of three other datasets. In particular, we present the approaches complementing one another and the gained insights, which will aid researchers in designing generic ridge detectors.

## 1 Introduction

Curvilinear structure enhancement and skeletonisation approaches are fundamental tools used in image processing. The enhancement process refers to a set of techniques that seek to improve the interpretability or perception of objects in the image for human observation or to provide better input for another image analysis task, such as ridge detection or segmentation. However, curvilinear skeletonisation helps to describe the shape of curvilinear structures as a single one pixel-wide path on pre-segmented (binary) images. Ridge detection is considered to be a special case of skeletonisation, where the extraction of the ridge is a free-segmentation process that is calculated directly from the grayscale image. Therefore, both skeletonisation and ridge detection processes can be referred to as line extraction or

ridge detection. Many enhancement and line extraction approaches have been discussed over the last several decades and extensive reports have been published [1, 2, 3, 4, 5]. Most line extractors were designed for a specific application, combining a line detection step with another line discrimination step using prior or contextual knowledge, such as the length or width of line segment (e.g., [6]). However, the images targeted for investigation in most applications require further refinement, such as noise reduction, in order to successfully extract lines. Therefore, researchers usually combine a prior enhancement step with the line extraction approach. The combination of these approaches is used in a vast number of applications in many domains, including but not limited to the following:

- Bioscience (e.g., fungal network detection [7]).
- Medicine (e.g., airway centreline extraction [8, 9], blood vessel extraction in the retina [10, 11, 12, 13, 14], cerebral vascular tree detection [11] and liver vascular network detection [15], virtual endoscopy [16]).
- Cosmetics industry (e.g., hair analysis [17, 18]).
- Engineering (e.g., fibre detection for industrial purposes [19]).
- Security (e.g., fingerprint detection [20]).

Extensive reports on these approaches have been published over the last several decades [1, 2, 3, 4, 5]. A wide variety of bespoke ridge detection approaches have been developed within related areas of morphological skeletonisation on binary images and skeletonisation of surface geometry, as well as more general ridge detection algorithms for imaging data [21]. However, it remains unclear as to which curvilinear structure enhancement methods best complement which ridge detection approaches, especially in generic (i.e., context-unaware) and highly varied real-world datasets [7]. A reliable analysis and a thorough evaluation of this relationship across large datasets present a significant challenge, both *statistically* (due to the varying high-dimensional search

---

HFA, CGW, SSA, BO are with the  
Department of Computer Science, Durham University, UK  
AK is with the  
Department of Anthropology, Durham University, UK  
HFA and SSA are with the  
Department of Information technology, Collage of Computer,  
Qassim University, Buraidah, KSA  
E-mail: boguslaw.obara@durham.ac.uk

space) and *computationally* (due to the large number of experiments required).

In general, parameters need to be specifically and carefully tuned for each application and, as a result, the parameters then do not generalise across scenarios [22]. In contrast, Convolutional Neural Networks (CNNs) [23], and more general machine learning approaches [24, 25], are excellent at handling such ill-defined problems. However, CNNs typically rely on large training datasets and may fail unpredictably, making them impractical for a number of applications.

In this paper, we study the relationship between curvilinear structure enhancement and ridge detection approaches, with particular emphasis on the following areas:

- **Impact:** Although new ridge detection approaches are continually emerging, it is difficult to understand whether and how the field is generally progressing, especially given that recent methods often rely on a separate pre-processing enhancement stage [7, 14].
- **Relationship:** It is not known which combinations of enhancement and ridge detection approaches work best together, and with which types of image data.
- **Robustness:** Some approaches may be able to achieve excellent results on certain types of data or particular curve-like patterns, but they may also exhibit unpredictable behaviour in other cases, such as in junction regions or in the presence of background inhomogeneities.
- **Parameters:** Some methods have large (e.g., 10+ dimensional) parameter spaces, but they may be more predictable, more intuitive (with flat and stable regions), and easier to tune, than other methods with smaller (e.g., 2-dimensional), but unintuitive parameter spaces, where the tuning of parameters may lead to substantial changes in the resulting image [26].

In this study, a large set of experiments tests each pair-wise combination of popular enhancement and ridge detection methods on two highly varied datasets and each combination is evaluated against their respective ground truths. Analysing these varied datasets allows us to collect important statistical information and to make inferences based on new insights about the effectiveness of individual methods and, more importantly, the overall relationship between enhancement and ridge detection. Quantitative and qualitative analyses are conducted for well-established and state-of-the-art methods (16 curvilinear enhancement methods, 3 segmentation methods, 6 skeletonisation methods, and 2 ridge detection methods that operate directly on image data).

*Contributions* The main contribution is a comprehensive benchmarking suite for assessing the generalisation of enhancement and ridge detection phases and, more importantly, for understanding the relationship between them. This is achieved by evaluating the statistical properties of the approaches on real-world datasets, for which we have created new ground truths. We also discuss the overall advancement in this field of research over the last few decades and we release our analysis and benchmarking tools, including the full source code, making it available for future use.

*Outline* In Section 2, we discuss the relevant literature covering related enhancement and ridge detection algorithms. In Section 3, we show how we choose our input images and create their ground truths. Section 4.1 evaluates the pair-wise method combinations on relevant datasets and introduces the metrics used along with the experimental results. In Section 5, we briefly discuss our results from Section 4.1 in relation to important cases. Section 6 provides details of our implementation and we conclude our study in Section 7.

## 2 Related Work

Literature in this section comes from many different fields which interchange terminology. For clarity, we refer to ‘skeletons’ as sets of line segments or curves, embedded typically in two or three dimensions (image space), representing local objects’ symmetries and their topological structure. We refer to ‘centrelines’ as thin rasterized connected lines or curves defined by the centres of curvilinear objects in binary and grayscale images (lines of bright pixels, ideally a binary image object which is 1-pixel thick). ‘Ridges’ are a set of curves whose points are local maxima on a surface defined by image intensities or as Lindeberg describes as where the intensity assumes a local maximum/minimum in the direction of the main principal curvature [27]. The term ‘curvilinear structures’ refers to wiry or tube-like structures in binary or grayscale images, with changing widths, curvature and orientation in different parts, that may crossover themselves or be occluded by other objects. ‘Enhancement’ refers to the process which transforms the image to increase the contrast between the curvilinear structures and their background, which may include other image objects and noise.

We categorise and discuss related literature in the subsections from the main research areas of structure enhancement, segmentation, skeletonisation, and ridge detection. Enhancement and segmentation approaches have significant overlap, as do the fields of ridge-detection

and skeletonisation, often differing only in application or in terms of how the shapes are represented.

## 2.1 Curvilinear Structure Enhancement

Curvilinear structures denote the presence of retinal vessels, cytoskeletal networks, fibres, etc. These examples often contain different curvilinear structures which are hard to capture, such as bends, crossing junctions, non-crossing junctions, etc. However, it is very difficult to accurately detect curvilinear structures due to the complexity of noise and inhomogeneous backgrounds. The principal aim of curvilinear structure enhancement is to modify the attributes of an image and to highlight the interesting curve-like structure in that image to improve its suitability for a given task or specific purpose (i.e., segmentation, ridge detection or centreline extraction). Most methods were proposed to improve a specific type or a region of vessels [28], whereas others focussed on noise reduction in the background [29] or were designed for a specific kind of image [30]. These methods used different techniques, ranging from integrating multi-scales, using multi-orientation kernels for regularisation, or to using different tensors to extract lines with variable widths and orientations. Therefore, these methods can be divided into several categories based on their mathematical principles.

### 2.1.1 Histogram-based Enhancement

Histogram equalisation is the most popular technique for improving image contrast. One of the earliest notable results in histogram equalisation was proposed by [31], who showed that selecting an appropriate clipping level for the histogram avoids undesired noise amplification and improves the flexibility in histogram remapping. Similarly, Contrast Limited Adaptive Histogram Equalisation (**CLAHE**) [32] and Constrained Local Histogram Equalisation (**CLHE**) [33] can enhance image details as histogram clipping techniques, which enhances image details while maintaining the overall image appearance. These approaches undertake various transformations of the same grey level at many spatial locations in the original image. In CLAHE, the image is divided into several regions and the histogram for each region is calculated individually. The same scenario occurs in CLHE and, in addition, the local grey level transformation is varied continuously by applying a concatenates condition. Recently, [34] proposed a dynamic fuzzy approach, which uses a triangular fuzzy membership function to provide better handling of inaccurate grey-level values than histogram clipping-based approaches, while preserving the mean image intensity. Another recent

work is [35] where they add naturalness preservation by using a bad illumination pass filter to locate badly illuminated areas in the image. Then, they proposed an adaptive logarithm transformation to add more brightness in poorly illuminated areas in the image. A major drawback of these methods is their sensitivity to noise due to the dependency on the image intensity.

### 2.1.2 Gaussian-based Enhancement

Gaussian kernels are amongst the most commonly used tools in image processing due to their properties, such as their isotropy, decomposability, integrability and differentiability [36]. Their fast computation makes them very convenient for low-level feature detection. Moreover, the derivatives of Gaussian kernels for enhancement add steerability which, when implemented as an orientation-based convolution kernel, can be used to calculate the response at different orientations. Freeman et al. [37, 36] provide the earliest results of using this in curvilinear structure enhancement. This work was improved in the Gaussian Steerable (**GS**) kernel [29] by introducing better edge recovery with greater spatial and angular accuracy than previous approaches. For anisotropic features, [38] introduced the first attempts at Anisotropic Gaussian kernels, developing a differentiation-based method that can detect local image conditions, such as width, roundness, and orientation by using the second partial derivative of anisotropic Gaussian kernels.

### 2.1.3 Mathematical Morphology-based Enhancement

Mathematical morphology operations bring a shape-based approach driven by set theory [39] to the image processing field. These operations are easy to implement and are indeed suitable for many shape-oriented problems. A large number of approaches have been proposed to enhance and to detect curvilinear structures based on different mathematical morphological transforms, such as [40, 41, 42]. In [40], Zana's popular Top Hat transform-based enhancement (**ZTH**) is proposed to identify image signals by preserving edges and prominent pixel information. It uses the Laplacian of Gaussian (LoG) for estimating an object's edges. In addition, the use of LoG also improves the speed of the implementation of this method. Similarly Sazak et al. [43] propose a recent multiscale bowler-hat transform which combines different structuring elements to detect innate features of vessel-like structures. The Ranking the Orientation Responses of Path Operators (**RORPO**) has been proposed to distinguish curvilinear objects from



blob-like and planar structures in images [41, 42]. However, current implementations incur a high computational cost and do not scale to large volumetric image datasets.

#### 2.1.4 Hessian-based Enhancement

Hessian-based approaches use second order local image intensity variations around the selected pixel and eigenvalue analysis of the Hessian matrix to describe different image features. The computation of the Hessian matrix requires an approximation of the second order partial derivatives. These approaches can be used to detect curved/tubular, plane-like, and blob-like features in 2D and 3D images [28]. In 1997, the first results using eigenvalues from a Hessian matrix analysis were obtained by [44] for vessel enhancement. One of the most popular Hessian-based approaches for enhancing curvilinear features was proposed by [28] and is known as Vesselness (**Vessel**). However, the main drawback of this approach are the very small responses for curvilinear features at junctions, due to similarly large eigenvalues. Later, [30] proposed another successful method, known as Neuriteness (**Neurit**). This is designed to enhance low contrast and highly inhomogeneous neurites in biomedical images, where a tuning parameter is used to modify the Hessian matrix calculation. In addition, they considered the maximum and minimum eigenvalues across the whole image to enhance bright, neurite-like structures. Most recently, [45] proposed a new measurement based on the Regularised Volume Ratio (**RVR**), which overcomes the deficiencies of using the Hessian matrix, namely non-uniformity, variation of eigenvalues with image intensity and non-uniformity of enhancement across scales. However, this approach is still sensitive at junctions and crossings, producing a lower response in these image regions.

#### 2.1.5 Phase Congruency Tensor-based Enhancement

The majority of Hessian-based approaches rely on image intensity, resulting in poor enhancement or detection of finer and lower intensity vessels. The Phase Congruency (PC) indicates the representation of the image in the frequency domain. The work by [46] proposed a contrast-independent approach using a Phase Congruency Tensor (PCT), which has previously been proven to be effective in detecting edges by [47] with advantages of insensitivity to intensity and noise variations. They introduced PCT versions of Hessian-based approaches, such as PCT-Vesselness (**PCTV**) [28] and PCT-Neuriteness (**PCTN**) [30]. However, a major drawback of this method is the complexity of its parameter space.

#### 2.1.6 Wavelet Transform-based Enhancement

Image transforms, such as the wavelet transform, are defined as tools for converting an image from one domain to another for easier identification of features than the spatial domain. In [48], the authors proposed an Isotropic Undecimated Wavelet Transform (**IUWT**), using a least-squares cubic B-spline fitting, followed by thresholding based on the wavelet coefficients. The number of thresholding levels has to be tailored to the specific application in order to achieve accuracy and computational efficiency. However, similar to Hessian-based approaches, wavelet transform-based approaches fail to enhance low-intensity and fine curvilinear structures.

### 2.2 Image Segmentation

Thresholding is a common technique for object segmentation, which can be categorised into two main approaches: global and local thresholding.

#### 2.2.1 Global Thresholding

In 1979, Otsu Thresholding (**OT**) was proposed, which exhaustively finds the optimal threshold to maximise inter-class variance [49]. Various Otsu-based approaches and extensions are reviewed in [50]. The main drawback of the (**OT**) is that this threshold only considers the intensity of each pixel point in isolation and ignores any relationships between the pixels. Another thresholding approach is Hysteresis Thresholding (**HT**), which applies dual thresholding followed by connected component analysis to prevent weak foreground areas from being merged into the background [51].

#### 2.2.2 Local Thresholding

This object segmentation approach involves finding the threshold value for individual pixels or regions of pixels, which depends on some local measures of spread, such as range, variance, or surface-fitting parameters of the pixel adjacency features. One of the most popular local methods is Adaptive Thresholding (**AT**) [52, 53], where the intensity values of the local neighbourhood of each pixel are examined. A comprehensive survey of image thresholding methods, reviewing all local techniques, can be found in [54]. However, the main drawback of thresholding is that choosing the appropriate thresholding values, suitable for all images in a dataset, remains a challenge.

## 2.3 Image Skeletonisation

A “skeleton” is an essential descriptor for shapes in many applications. Blum [55] proposed a medial axis definition as a simulated grassfire transform process and established the foundation of “skeletonisation”. This medial axis can be described as planes/axes of symmetry with lower dimensionality. While there is no explicit definition of a curve-skeleton, Cornea et al. [56] described some desirable properties of a curve-skeleton and its computation process. The availability of these properties are dependent on the application. These properties include: 1) the skeleton should be thin, 2) it should be one-pixel/voxel wide, 3) it should capture the shape of the object, 4) it should be centred between object boundaries, 5) it should be connected, 6) it can offer component-wise differentiation for the segments of the object, and 7) every inner boundary point should be visible from the skeleton. Furthermore, the skeleton computation process should be robust and insensitive to noise. The object can be reconstructed in a computationally efficient manner and follows a hierarchical scheme. A categorisation of the various approaches for skeletonisation was proposed by [5], which allows adequate abstraction of the principles for each approach. Similarly, there are approaches that follow or perturb the object boundary, such as computing a distance transform, distance field or using a gradient vector field [16, 57, 58, 59]. Siddiqi et al. [60, 61] proposed a Shock Graph-based approach (SG), a Hamiltonian formation of curve continuous transformation, computing the outward flux of the vector field of the underlying system using the Hamilton-Jacobi equation. Skeletons are located at singularities of this flux field. Furthermore, the topology preservation constraints of digital grids have been utilised to enhance the robustness of computed skeletons. [62] introduced the Integer Medial Axis skeleton using an improved form of a linear-time algorithm for Euclidean distance transforms and proved a set of fundamental properties of the IMA skeleton. They then compared these properties to those of the Centres of Maximal Disks skeleton. Telea [63] proposed a robust Augmented Fast Marching Method (AFMM) to compute a parametrised boundary location for every pixel during the boundary evolution. The parameter field is thresholded to produce the skeleton from boundary features.

### 2.3.1 Propagation-based Approaches

These skeletonisation approaches work according to Blum’s grassfire propagation strategy. The first approach repeatedly employs Morphological Thinning (MT) [64]

by shifting boundary pixels by one pixel without disturbing the shape’s local connectivity until it reaches single pixel width. In 1984, [65] proposed a Zhang Standard Thinning (ZST), followed by [66] who improved the results in terms of line connectivity. [67] overcomes the connectivity issues identified in [65] with a hybrid approach that produces thinner results. Vincent et al. proposed Morphological Skeletonisation [68] as an advanced approach to image analysis. Later, the “maximal balls” concept was introduced with a weighted distance transform [69], and further developed by [70] with larger weighted distances. In 2000, [71] proposed a fuzzy morphological skeletal procedure for 2D and 3D objects and [11] used a fuzzy ordered region-growing approach in skeletonisation.

### 2.3.2 Geometric-based Approaches

Many approaches have been proposed based on the geometric properties of Blum’s medial symmetry axis. One of the most popular geometric-based approaches is Voronoi Diagram skeletonisation (VD) [72, 73], producing a large number of false skeletal segments with a low significance measure in the skeleton’s hierarchy, which should be pruned. Another geometric approach iteratively smoothes the surface geometry [74], contracting the surface into a skeleton. The amount of smoothing controls the topology and size of features captured in the skeleton, but it is not homotopic or centred.

### 2.3.3 Learning-based Approaches

These recent approaches are more tailored to detecting skeletons from complex scenes, such as natural images [75, 76, 77]. By employing large numbers of parameters on large training datasets, [75] reform the skeleton extraction problem, positing that it should be solved as a regression problem. As a result, their solution is more generic and well-suited to solve other problems. More recently, [76, 77] proposed a “multi-clusters” of Multiple Instance Subspace Learning (MISL) model based on a divide-and-conquer strategy for problem solving.

They create object skeleton ground truths at different scales and orientations, where the final output is gathered from symmetry results in subspaces. Training classifiers in subsets improves the discrimination of the symmetry detector, which leads to performance improvements. This approach is not limited to skeletonisation, but is generic to some detection tasks in computer vision. Similarly, recent work uses the popular U-Net deep neural network segmentation method to estimate the probability of pixels being on the centreline.

This gives thick outputs which need to be further post-processed using traditional methods [78]. Although the learning-based methods can obtain good results, attaining ground truths can be expensive in terms of both data acquisition and the training process.

## 2.4 Ridge Detection

Ridges are shape descriptors that provide an approximate symmetrical axis of the curvilinear object [27, 79]. They are considered as directional local extrema [80], where Lindeberg proposes an automatic selection criteria of normalised derivatives of multi-scale Gaussian convolution, achieving Edge Detection and Ridge Detection with Automatic Scale Selection (**EDRDASS**) [27]. Another popular Unbiased Ridge Detector of Curvilinear Structures (**URDCS**) was proposed by Steger [81], convolving Gaussian kernels and searching for high curvature without bias in asymmetrical lines. A successful approach for ridge detection utilises steerable Gaussian functions with an optimised Canny operator and improved orientation selectivity [82].

### 2.4.1 Hybrid Ridge Detectors

Recently, most ridge detectors comprise a pre-ridge detection phase that enhances ridge features before fitting [83]. The Anisotropic Gaussian Kernel-based ridge detector (**AGK**) [7] has extended the work by [38], focussing on ridges and valleys through introducing a hybrid multi-scale ridge detector that combines enhancement and ridge detection. This method allows detection of curvilinear structures with different widths, using anisotropic Gaussian smoothing phases to detect curvilinear features. There are dependencies between enhancement and detection and, therefore, we are particularly interested in the enhancement phase of the work done by [7], where excellent performance with regard to sensitivity to crossings and junctions, as well as providing better characterisation of blob-like artifacts was achieved. Another issue is that using the full pipeline of **AGK** in our study caused blurring or diminished the target curvilinear structure in the image. A novel hand-crafted feature Scale and Curvature Invariant Ridge Detector (**SCIRD**) has been proposed to achieve multiple invariances when segmenting tortuous and fragmented structures [14], using an enhancement step followed by a ridge detection step. This ridge detector does not guarantee the production of “thin” ridges (of approximately 1 pixel or voxel in thickness). Therefore, we can not use the full pipeline of this approach, but we can utilise its enhancement step when comparing enhancement methods. In this paper, we analyse

the relationship between well-known enhancement and skeletonisation/ridge detection approaches. We discuss the underlying principles, challenges, advantages and drawbacks and summarise all candidate approaches for this analysis in Table 1, based on our criteria in Section 4.1.1. We show how recent methods fare against older ones and whether there is general improvement or overfitting on certain datasets.

## 3 Materials

Two substantially different biomedical datasets were chosen for this analysis: Digital Retinal Images for Vessel Extraction (DRIVE) [6, 84] consists of 40 images, along with manual segmentation of the vessels, and Ghent University Fungal Images (GUF1-1) [7] is a dataset of 100 images, which were extracted from fungi grown in vitro. Each of the images has a resolution of  $300 \times 300$  pixels and comes with their respective hand-labeled segmentation solution.

These datasets were selected based on the following four criteria: 1) their popularity, 2) containing a sufficient number of images, 3) having various types of biases, such as containing varying patterns, different thicknesses of the vascular structures and background noise, and 4) potentially useful for future research for both medical and biological purposes. Particularly, in DRIVE, there are problems relating to intensity inhomogeneities, and that it is a well-known dataset which has been used to test and to evaluate curve-like detectors. In contrast, GUF1-1 contains an assortment of ridges of different widths, junctions and densities.

### 3.1 Input Images and Ground Truth

The skeleton ground truths of these datasets are not available (DRIVE), or are not always thin (i.e, with 1 pixel width) (GUF1-1). Therefore, we created new ground truths to ensure availability and the required thickness. The ground truths have been established independently by two experts and the mean Hausdorff Distance (MHD) calculated between the two sets in order to ensure the validity of the proposed ground truth. The mean of MHD is 2.22 pixels for DRIVE and 0.82 pixels for GUF1-1, where the standard deviation of HD is 0.82 pixels for DRIVE and 0.86 pixels for GUF1-1. Figure 1 shows examples for each dataset and the corresponding manual skeleton that we created. In addition, the ground truths of samples from other high resolution datasets were created, including DR HAGIS [86], HRF [87] and CrackForest [88] in order to further evaluate the findings of our study for different data scenarios.

Table 1: A selection of well-known previously published methods for curve-like features enhancement, thresholding, skeletonisation and ridge detection, sorted by their publication year. M = Matlab, C = C/C++, J = JAVA.

No.	Category	Method	Pub.	Year	Implementation	Highlight
1	Enhancement	GS [29]	ECCV	1992	M	Multi-scale multi-orientation kernels approximated by linear interpolation
2		CLAHE [32]	GG IV	1994	M	Contrast-limited adaptive histogram equalization
3		Vessel. [28]	MICCAI	1998	M	Vessel enhancement filter using multi-scale second order local structure of Hessian matrix of the image
4		LTV [44]	MICCAI	1998	M	Multi-scale Hessian-based approach that takes the maximum of single-scale response
5		CLHE [33]	CVIU	1999	M	Constrained Local Histogram Equalization in variational form
6		PCT [85]	Videre	1999	M	Phase congruency calculation using Gabor wavelets
7		ZTH [40]	ITIP	2001	M	Mathematical morphology and curvature evaluation for the detection of vessels
8		Neurit. [30]	Cytometry Part A	2004	M	A semi-automatic neurite tracing technique using second-order image derivatives of the Gaussian kernel and a modified Hessian matrix
9		BPDFHE [34]	ITCE	2010	M	Brightness Preserving Dynamic Fuzzy Histogram Equalization
10		PCT Vessel. [46]	ITIP	2012	M	Contrast-independent approach to vesselness detection based on phase congruency tensor
11		PCT Neurit. [46]	ITIP	2012	M	Contrast-independent approach to neurite tracing based on phase congruency tensor
12		IUWT [48]	PLoS one	2012	M	Thresholding based on coefficients of B-spline fitting Isotropic Undecimated Wavelet Transform
13		AGK [7]	SP	2015	M	Multiscale non-steerable Anisotropic Gaussian kernels (enhancement part only)
14		SCIRD [14]	MICCAI	2015	M	Scale and curvature in-variant, and relaxes shape assumptions to achieve enhancement of target image structures (enhancement part only)
15		RVR [45]	ITMI	2016	M	Regularised volume ratio applied on eigenvalues of Hessian matrix of the image
16		RORPO [41, 42]	ITPAMI	2018	C	Ranking the Orientation Responses of Path Operators aiming to detect curve-like and blob-like objects
1	Segmentation	OT [49]	ITSMC	1979	M	Automatic threshold selection to maximise separability of the resultant classes in gray levels
2		AT [53]	ITPAMI	2003	M	Adaptive local thresholding based on a multi-thresholding verification scheme
3		HT [51]	AP	2012	M	Hysteresis thresholding for edge detection using two thresholds
1	Skeletonisation	MT [64]	ITASSP	1986	M	Morphological set operations to represent and encode a discrete binary image
2		ZST [66]	PRL	1988	M	A modified version of Zhang and Suen thinning algorithm which avoid serious shrinking and line connectivity
3		MS [68]	ITIP	1993	M	Hybrid grayscale morphological skeleton reconstruction
4		SG [60, 61]	CCV	1999	M	Skeletonisation derived from the shocks (singularities) of a curve evolution process
5		AFMM [63]	PSDV	2002	M+C	A robust augmented Fast Marching Method to compute skeleton from boundary features.
6		VD [72]	VDSE	2006	M+C	Employing Delaunay triangulation and Voronoi diagram in order to extract skeletons
1	Ridge detection	EDRDASS [27]	CVPR	1996	M	An automatic selection of scale levels when detecting one-dimensional images features
2		URDCS [81]	TPAMI	1998	J	Lines Extraction using Gaussian masks with removing bias which is induced by asymmetrical lines to estimate unbiased derivatives

## 4 Method

### 4.1 Quantitative Analysis of Enhancement and Ridge Detection Combinations

#### 4.1.1 Algorithms

A chosen selection of enhancement, skeletonisation and ridge detection approaches are discussed in Sections 2.1,

2.3 and 2.4, respectively. Our criteria for choosing these approaches were: (1) the approach is well-known in the research field, (2) the approach is reasonably easy to implement, and (3) due to high computational cost of this experiment, the underlying algorithm is reasonably efficient, not causing a substantial bottleneck in the overall framework performance.

We evaluate these approaches against real medical and biological image datasets. This allows us to identify the relationship between these two fields and to infer a generalised evaluation. In particular, we optimise enhancement and detection parameters in a large non-differentiable search space, minimised using a genetic algorithm [89, 90] on a large high-memory server. In this comparison, we have evaluated combinations of 27 approaches in total (16 enhancements, 3 segmentations, 6 skeletonisations and 2 ridge detection approaches) (see Table 1 for more details). We selected 16 enhancement approaches from different categories (c.f., subsection 2.1) and combined them with skeletonisation/ ridge detection approaches, with the goal of extracting a single pixel ‘thin’ ridge as the final result. The manual skeleton annotation, i.e., Ground Truth (GT), is the baseline model for comparison. In order to evaluate each combination of enhancement and ridge detector, we conducted 152 large experiments to examine different parameter spaces (c.f., Table 2). Some of the enhancement approaches included in this comparison are considered as preprocessing stages in hybrid ridge detection approaches, where we have extracted the enhancement phase as a separate approach. In these hybrid approaches, the ridge identification step depends on the enhancement step, e.g., **AGK** [7] and **SCIRD** [14],

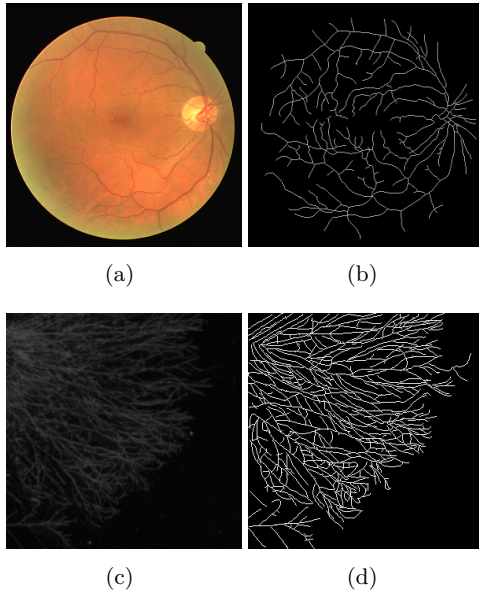


Fig. 1: Samples images alongside the ridge ground truths of used datasets. (a) An example of retinal images from the DRIVE dataset, (b) The corresponding skeleton ground truth of (a), (c) An example of fungal images from the GUF1-1 dataset, (d) The corresponding ridge GT of (c).



therefore we could not compare these ridge detectors with other approaches due to these dependencies.

#### 4.1.2 Evaluation Measures

Visual tasks are often based on the evaluation of similarities between image-objects represented in an appropriate feature space. Several metrics exist to assess the similarities between shape geometry. For binary images, similarities are assessed by considering a set of black pixels as an object, and the remaining pixels as the background. In [91], they compared several measures of correspondence between binary images. The comparison shows that distance-based measures work better for binary images than measures based on set memberships. In 1992, Baddeley [92] proposed a new error metric for binary images that calculated distances from every point of a two dimensional image space to the nearest object pixel. Huttenlocher et al. [93] have proposed a robust and tolerant object similarity measure based on the Hausdorff Distance (HD) and it has since become a generally accepted standard metric for evaluating distances between two subsets of metric space. We use similarity metrics to compare two images and we define the two-dimensional digital images  $A$  and  $B$  as a discrete function defined in a lattice domain  $D$  of size  $N \times N$ , taking values in the set of grey levels  $0, 1, 2, \dots, 255$ . Here, without losing generality, we consider an image  $A$  as a set of pixels  $A_{ij}$  and an image  $B$  as a set of pixels  $B_{ij}$ , where every pixel is defined by its spatial coordinates  $(i, j)$  and grey value  $a_{ij}$  and  $b_{ij}$ . The HD is defined mathematically as the inter-pixel distance between the images  $A$  and  $B$ , as follows:

$$h_{HD}(A, B) = \max_i j(d(A_{ij}, B), d(B_{ij}, A)), \quad (1)$$

where the directed Euclidean distance  $d(A, B)$  is defined as:

$$d(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|. \quad (2)$$

This provides a worst-case measure of all pairs of distances, serving in many applications, such as object recognition and object matching [94, 95], and has been suggested for object matching by [93] and other researchers [96, 97, 98]. The HD is very sensitive to outliers. A single outlier can perturb the distance greatly, even though the two objects might be very similar. This sensitivity can be seen as an advantage in the application of otherwise highly similar subsets of metric spaces.

A popular adaptation of Huttenlocher's work is called the Modified Hausdorff Distance (MHD) [99], which has many applications in assessing similarities in image processing and computer vision [100, 101, 102, 103]. It is defined as:

$$h_{MHD}(A, B) = \max(md(A, B), md(B, A)), \quad (3)$$

where the directed Euclidean distance for MHD  $md(A, B)$  is defined as:

$$md(A, B) = \frac{1}{|A|} \sum_{a_n \in A} \min_{b_n \in B} \|a - b\|. \quad (4)$$

and  $\min_{a_n \in A} \|a - b\|$  represents the minimum distance at point  $a$  to the point  $b$  and  $n$  is the number of pixels in image  $A$  or  $B$ . [99] investigated 24 forms of different Hausdorff distance measures and indicated that the MHD measure gives the best performance in the application of object matching.

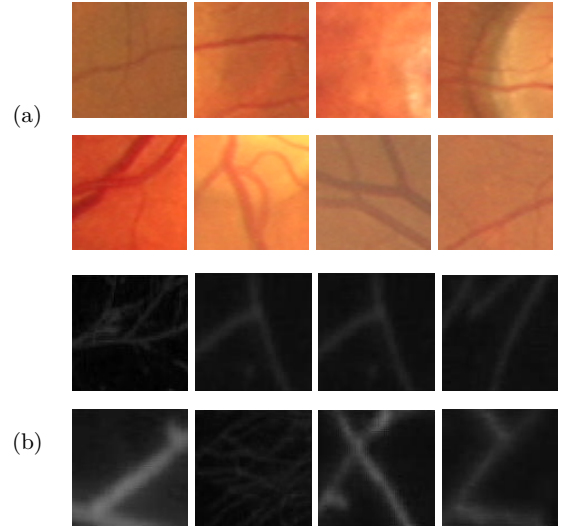


Fig. 2: Examples images of used datasets in quantitative analysis of relationship between enhancement and ridge detection. (a) A selection of cropped retinal images form DRIVE dataset and (b) A selection of cropped fungal images from GUF1-1 dataset.

#### 4.1.3 Parameters Search

Given the number of experiments required by this study, we performed initial calculations based on a small subset of the original dataset. We have selected ten small cropped image regions ( $50 \times 50$  pixels), representing a variety of different and challenging cases; vessels of varying thickness, inhomogeneity, weak/blurred edges, crossover, and bright vessels against dark vessels, as shown in Figure 2. This selection is used to refine our search space down to a subset of approach combinations, from which we conduct experiments on the full dataset at the original resolution. We evaluate ridges extracted by different approaches on different datasets (as described in Subsection 3) using the evaluation criteria defined in the following section.

Table 2: The total number of parameters of enhancement and ridge detectors combinations. Light blue coloured cells indicate less parameters are required. The approaches are sorted based on the number of required parameters.

Enh/Seg Skeleton	AT	OT	CLAHE	Neu	CLHE	TH	IUWT	HT	GS	RVR	RORPO	Vessel	LTV	PCTV	PCTN	PCT	BPDFHE	AGK	SCIRD
AFMM	2	2	2	2	3	3	3	4	4	5	5	8	9	9	10	10	11	14	13
ZST	2	2	2	2	3	3	3	4	4	5	5	8	9	9	10	10	11	14	13
MT	3	3	3	3	4	4	4	5	5	6	6	9	10	10	11	11	12	15	14
EDRDASS	3	3	3	3	4	4	4	5	5	6	6	9	10	10	11	11	12	15	14
MS	3	3	3	3	4	4	4	5	5	6	6	9	10	10	11	11	12	15	14
SG	4	4	4	4	5	5	5	6	6	7	7	10	11	11	12	12	13	16	15
VD	5	5	5	5	6	6	6	7	8	8	7	11	12	12	13	13	14	17	16
URDCS	5	5	5	5	6	6	6	7	7	8	8	11	12	12	13	13	14	17	16

#### 4.1.4 Evaluation Criteria

We calculated the mean of all genetic-driven experiment of each combination to show how close it can achieve to the corresponding GT. In addition, the standard deviation (std) shows the variation in the results of the combination on the same dataset with different curve-like patterns and different parameters. We are seeking to find pairs of approaches that generalise well to a variety of scenarios and, therefore, use both the Grand Mean (GM) of the MHD and the Root Mean Square (RMS) of the MHD in our assessment. Moreover, in order to investigate the correlation between the results statistically and to measure the overall reproducibility of each row/column, we compared the results of each row/column using Intraclass Correlation Coefficient (ICC1) [104] to measure within the same subject (images). An intraclass correlation is a measure of correlation between raters and provides an indication of the reliability of ratings (enhancement and skeletonisation) [105].

Furthermore, we test whether the assumption of the effectiveness of each approach from a certain field is valid in another field by computing the p-value. A significant p-value means that the values were different. Thus, we apply a test called Likelihood Ratio Test (LRT), which compared the two hypotheses. The null hypothesis (H0) tested whether the values in the column OT are the same. The alternative hypothesis (H1) implies that the values are not equal. LRT is corrected by False Discovery Rate correction (FDR) using the Benjamini-Hochberg Procedure [106], which is a procedure to control the FDR by reducing it and to adjust the p-value for multiple comparisons.

## 4.2 Quantitative and Qualitative Analysis of Different Curvilinear Structure Patterns from Other Datasets

In this section, we assess our key findings on other recent datasets, in effect evaluating the generalisation performance of our study. These datasets are of higher

resolution (and quality) and serve more general-purpose usages (comparing the quality of the combined approaches, sensitivity to input parameters, and the data scenario for which they are best suited). We compared the selected approaches with previous quantitative data, with an example of enhancement approaches that achieve average results on both datasets in the previous analysis experiment, and the most recent enhancement approaches that overcome most of curvilinear structures enhancement deficiencies. All chosen enhancement approaches have been combined with the **URDCS** skeletonisation algorithm and used the non-differentiable optimisation strategy (as in Section 4.1) to find the global minimum representing the best possible solution for certain patterns.

#### 4.2.1 Materials

All combinations have been examined with samples from many datasets, including DR HAGIS [86], HRF [87] and CrackForest [88]. We assessed their respective strengths and weaknesses, using different patches in  $200 \times 200$  pixel regions from each group in these datasets, and we manually created new ground truths of the skeletons for each image used.

#### 4.2.2 Quantitative Analysis on Unseen Datasets

We use the same mean of the MHD that has been used in Section 4.1 to measure the overall performance of combinations across different datasets, as shown in Table 7. The results reflect our findings in the larger experiments, shown in Tables 9 - 12, where the **AGK-URDCS** combination is shown to outperform other approaches.

## 4.3 Generalised Quantitative and Qualitative Analysis

To infer a generalised evaluation, we simulate the real world images by comparing the recommended approach with other interesting approaches, showing the results



Table 3: Comparing the mean of the MHD between the GT and the extracted ridges, for most interested enhancement and ridge detector combinations using images from DRIVE. The last four rows and columns show the GM, ICC1, LRT and BH(p-value) across all enhancement and ridge detection methods respectively.

Enh/Seg	PCT	AGK	SCIRD	RVR	GM	ICC1	p-value	BH(p-value)
Skeleton								
MT	5.26	3.65	4.73	3.41	2.52	0.56	$4.33 \cdot 10^{-2}$	$3.46 \cdot 10^{-2}$
ZST	6.13	2.88	5.37	3.16	2.56	0.32	$2.1 \cdot 10^{-2}$	$4.41 \cdot 10^{-2}$
MS	5.00	2.75	4.90	3.21	2.34	0.18	0.14	0.16
EDRDASS	5.38	6.02	5.11	6.61	3.45	0.69	0.19	0.19
URDCS	4.06	1.98	5.50	3.41	2.25	0.50	0.13	0.16
SG	6.40	7.98	6.96	5.77	3.81	0.23	0.12	0.16
VD	9.19	6.69	5.63	5.80	3.94	0.24	$2.21 \cdot 10^{-2}$	$4.41 \cdot 10^{-2}$
AFMM	4.61	3.16	5.21	4.01	2.53	0.62	$2.15 \cdot 10^{-2}$	$4.41 \cdot 10^{-2}$
GM	4.25	3.16	3.95	3.22				
ICC1	0.21	0.37	$9.04 \cdot 10^{-2}$	0				
p-value	0.26	$8.83 \cdot 10^{-2}$	$6.08 \cdot 10^{-4}$	$3.61 \cdot 10^{-4}$				
BH(p-value)	0.27	0.15	$2.84 \cdot 10^{-3}$	$2.29 \cdot 10^{-3}$				

Table 4: Comparing the mean of the MHD between the GT and the extracted ridges, for the most interested enhancement and ridge detector combinations using images from GUF1-1. The last four rows and columns show the GM, ICC1, LRT and BH(p-value) across all enhancement and ridge detection methods respectively.

Enh/Seg	PCT	AGK	SCIRD	RVR	GM	ICC1	p-value	BH(p-value)
Skeleton								
MT	5.78	5.57	5.64	6.32	3.34	$5.33 \cdot 10^{-2}$	$4.34 \cdot 10^{-4}$	$1.74 \cdot 10^{-3}$
ZST	6.02	6.52	6.03	7.61	3.80	0.43	$7.3 \cdot 10^{-3}$	$1.95 \cdot 10^{-2}$
MS	5.57	5.73	5.93	6.47	3.42	0.23	$1.67 \cdot 10^{-2}$	$2.67 \cdot 10^{-2}$
EDRDASS	6.02	6.84	6.03	6.90	3.76	0.50	$1.2 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$
URDCS	4.88	5.31	5.22	5.40	3.13	0.54	0.26	0.3
SG	7.01	7.16	7.10	6.44	4.13	0.32	$4.57 \cdot 10^{-2}$	$6.1 \cdot 10^{-2}$
VD	9.19	6.79	6.06	6.48	4.16	0.50	$4.57 \cdot 10^{-2}$	$6.1 \cdot 10^{-2}$
AFMM	5.31	5.71	5.35	6.68	3.30	$4.63 \cdot 10^{-28}$	$2.83 \cdot 10^{-7}$	$2.26 \cdot 10^{-6}$
GM	4.61	4.55	4.3	4.76				
ICC1	0.36	0.21	0	0				
p-value	0.3	$8.69 \cdot 10^{-2}$	$5.85 \cdot 10^{-4}$	$6.8 \cdot 10^{-4}$				
BH(p-value)	0.3	$8.69 \cdot 10^{-2}$	$5.85 \cdot 10^{-4}$	$6.8 \cdot 10^{-4}$				

Table 5: Comparing the standard deviation of the MHD between the ground truth and the extracted ridges, for the most interested enhancement and ridge detector combinations using images from DRIVE. The last row and column show the Root Mean Square (RMS) across all enhancement and ridge detection methods respectively.

Enh/Seg	PCT	AGK	SCIRD	RVR	RMS
Skeleton					
MT	1.66	1.73	1.07	2.50	1.81
ZST	2.28	1.63	2.37	0.00	1.84
MS	1.37	1.91	1.43	2.53	1.87
EDRDASS	1.34	1.46	1.38	1.36	1.38
URDCS	1.48	1.46	1.38	2.01	1.60
SG	1.34	1.25	1.26	1.69	1.40
VD	0.70	1.78	1.13	2.09	1.52
AFMM	1.61	1.83	1.34	2.92	
RMS	1.64	1.76	1.57	2.21	

of full-size images from DRIVE and GUF1-1 datasets separately and mixed, respectively, as shown in Table 8.

## 5 Results and Discussion

In order to ensure a fair comparison in Section 4.1, we systematically optimised the parameters for all tested enhancement/ridge detector combinations.

Based on the obtained parameters, we presented the results of our experiments based on two different datasets, with varying curvilinear structure patterns, resolutions and background difficulties.

Table 6: Comparing the standard deviation of the MHD between the ground truth and the extracted ridges, for the most interested enhancement and ridge detector combinations using images from GUF1-1. The last row and column show the Root Mean Square (RMS) across all enhancement and ridge detection methods respectively.

Enh/Seg	PCT	AGK	SCIRD	RVR	RMS
Skeleton					
MT	1.23	0.90	0.96	0.80	0.98
ZST	1.75	1.90	1.34	2.63	1.96
MS	0.99	1.11	0.88	1.20	1.05
EDRDASS	0.93	1.22	1.08	1.50	1.20
URDCS	0.88	0.89	1.12	1.02	0.98
SG	1.74	2.02	1.39	1.18	1.62
VD	0.70	1.50	1.07	1.46	1.23
AFMM	1.08	1.42	0.75	1.29	1.16
RMS	1.3	1.52	1.17	1.58	1.40

In addition, we considered various metrics to summarise the algorithmic performances in order to present an overall evaluation of the relationship between these two fields:

- Tables 3 and 4 compare the most interesting enhancement and ridge detectors for images from both datasets, DRIVE and GUF1-1, using the mean of MHD<sup>a</sup>. Lighter coloured cells indicate a smaller mean of MHD scores, which is similar to their ground truth, whereas dark blue indicates a higher mean of MHD scores. In addition, the last three rows and columns show the the overall evaluation of all methods. The last two rows in both tables show the overall evaluation of an enhancement approach across all ridge detectors. The last two columns reflect the overall evaluation of behaviour for each ridge detector across all pre-ridge detection approaches. One row shows the resulting p-values from LRT and the last row shows the adjusted p-values for multiple comparisons using False Discovery Rate correction (FDR) using the Benjamini-Hochberg Procedure. Both tables show ICC1 performed on both datasets to determine whether the results are correlated and reproducible.
- Tables 5 and 6 show the stability of output measures of the most promising approaches for selected images from both datasets, DRIVE and GUF1-1, using the standard deviation of MHD<sup>b</sup>. In other words, it shows how sensitive the combinations are to parameter variations and how robust they are for extracting a wide range of curve-like patterns. Lighter coloured cells indicate a smaller standard

<sup>a</sup> For the comparison of all enhancement and ridge detectors for images from both datasets, DRIVE and GUF1-1, the reader is advised to refer to Tables 9 and 10.

<sup>b</sup> To see the stability of output measures of all enhancement and ridge detectors for images on both datasets, DRIVE and GUF1-1, the reader is advised to refer to Tables 11 and 12.

deviation of MHD scores, which indicates robustness across different curve patterns. Light blue indicates a higher standard deviation of MHD scores. The last row in both tables show the overall evaluation of stability of an enhancement approach across all ridge detectors. The last column reflects the overall evaluation of stability of results for ridge detectors across all pre-ridge detection approaches.

Given the results, and taking a closer look at the combination scores across image samples shown in Figure 3 (a), we conclude that **AGK** comes closest to the ground truth across the DRIVE sample images. **SCIRD** performs better on GUF1-1 than DRIVE images. In Tables 3 and 4, it is notable that the results for **AGK** are more correlated than for **SCIRD**, whilst **SCIRD** is showing significant adjusted scores ( $BH(p\text{-value}) < 0.05$ ). The performance stability shown in Tables 5 and 6 for **SCIRD** has smaller RMS scores, which means that it is more stable than **AGK** when it is used across different data. We observed that local kernel adaptation and scale-based approximation can improve the results of image enhancement. However, these results reinforce the view that enhancement algorithms are sensitive to the thickness of curve-like structures, as well as the amount, type of noise and background illumination.

For morphological-based approaches, **ZTH** achieves good results and outperforms others from the same class with notable stability on both datasets, due to its insensitivity to background noise. For comparison, **RORPO**'s results are more correlated and statistically significant in GUF1-1 ( $BH(p\text{-value}) < 0.05$ ), where the backgrounds have less noise.

Given the fact that **RVR** excels at junctions when compared to other Hessian-based approaches, such as **Vessel** and **Neu**, **RVR** outperforms the other approaches ( $BH(p\text{-value}) < 0.05$ ) across both datasets. On the other hand, we find that **RVR** results are less correlated. This can be clearly seen in Figure 3 (e) when it is combined with **URDCS** and compared with other ridge detectors from the same class, such as **EDRDASS** in Figure 3 (n). In **URDCS** approach [81], the line is precisely extracted by solving the problem in the bias of line detection.

Notwithstanding the fact that parameters of **PCT** approaches are hard to tune, Tables 3-6 show that the parameters are both correlated and stable, compared to all other enhancement approaches, and they well-capture thin curves. We can see an example of its stability by comparing its performance combined with **URDCS** and **EDRDASS** ridge detectors in Figure 3 (h) and (o).

Among Histogram Equalisation-based techniques, they generally have average scores and are less robust. However, **BPDFHE** shows a high correlation between multiple comparisons. It works much better on thick curves than thin ones as can be seen in Figure 3 (i).

Since **GS** uses the steerability property of Gaussian, it shows a high correlation across both datasets and more stability on less noisy images, in this case the GUF1-1 dataset. When combined with the **URDCS** line detector, shown in Figure 3 (k), it achieves better results on thick curves in the DRIVE dataset than on thin curves in the GUF1-1 dataset.

The individual combination of **AGK** and **URDCS** has performed best over all significant experiment combinations across all datasets.

Qualitatively, we can see in Table 7 that the proposed combination performs well in the pre-ridge detection processes, but may still result in disconnectivity at the junctions in cases of intensity inhomogeneities.

We believe that the main advantage of **AGK** and **SCIRD** over other enhancement approaches is their ability to reduce background noise, allowing the ridge detector to achieve more accurate results. The effect of noise and intensity inhomogeneities in the background with **PCT** and **RVR** decreases the effectiveness of curvilinear structure enhancement, influencing the final result of the ridge detectors.

The results in Table 7 show false positives and false negatives in a variety of cases.

In particular, the limitations within the enhancement phase causes failure to detect the underlying curvilinear structures in cases of intensity inhomogeneity or noise. The ridge detection phase (primarily **URDCS**) often fails given the amount of variation in the width of the curve-like structures.

Both quantitative and qualitative results (c.f. Table 8) show the advantage of our proposed combination over other candidates. The **AGK-URDCS** combination is least sensitive to noise and inhomogeneous backgrounds across the datasets, whilst being able to handle a high degree of variation in terms of network complexity and complex backgrounds. Additionally, the proposed combination robustly extracts thin one-pixel wide skeletons with few complete fail cases.

We also noticed that the proposed combination typically outputs well-connected skeletons, due to the anisotropic local kernel-based enhancement phase targeted at junction detection.

Combining the **URDCS** skeleton method with the **AGK** enhancement method achieved better results for thin vessels compared with **SCIRD**, as shown in the GUF1-1 dataset. Consequently, we recommend this combination over others.

In summary, we have found that the enhancement phase is a critical component for ridge detection. A good enhancement approach has to achieve the following: i) insensitivity to noise and intensity non-uniformities (c.f., for example, Table 7 (rows 2-3) and Table 8 (rows 1-2)), which should focus on eliminating the effect of background noise or intensity inhomogeneity in or outside the curvilinear structure, ii) robustness at junctions/crossings, where the method should not rely on distinct or clearly defined background regions at either side of the curve (see for example Table 7 (rows 4-7) and Table 8 (rows 3-4)), iii) variable vascular morphology, where the solution should support fitting elongated shapes in order to characterise the local manifestations of certain global curvilinear structures (see for example Figure 2 and Table 7), and iv) variable vascular widths: (c.f., for example Table 7 (rows 2,6)) where high variations in scale lead to false positives or false negatives.

## 6 Availability and Implementation

- The software has been implemented in MATLAB R2017a and made available at:  
<https://github.com/Haifafh/Enhancement-Ridge-relationship>.
- The manually corrected skeleton datasets for DRIVE and GUF1-1 are also available at:  
<https://github.com/Haifafh/skeleton-GT-DRIVE-GUF1-dataset>.

## 7 Conclusion

In this study, we have quantitatively investigated the impact of pairwise combinations of enhancement and ridge detection methods on real-world medical and biological datasets. We manually created new ground truth skeletal datasets and extensively evaluated parameter spaces using large genetic-driven experiments on a high-memory server.

Overall, we found that traditional ridge detectors, such as the URDCS ridge detector [81], work well with more recent enhancement methods, which utilise multi-scale customised kernels. Furthermore, we have shown quantitatively that curvilinear enhancement remains a critical component in the design of generic ridge detectors. Future work and development of this study is to extend our evaluation to 3D images. Another branch of further research is to use our framework to investigate specific biological or medical tissue types. Moreover, the evaluation can be extended to assess the skeleton based on measuring the connectivity of the skeleton network.

## 8 Acknowledgment

Haifa Alhasson and Shuaa Alharbi are supported by the Saudi Arabian Ministry of Education Doctoral Scholarship and Qassim University in Saudi Arabia.

## References

1. D. C. Wang, A. H. Vagnucci, and C. Li, “Digital image enhancement: a survey,” *Computer Vision, Graphics, and Image Processing*, vol. 24, no. 3, pp. 363–381, 1983.
2. M. S. Miri and A. Mahloojifar, “A comparison study to evaluate retinal image enhancement techniques,” in *IEEE International Conference on Signal and Image Processing Applications*, (Kuala Lumpur, Malaysia), pp. 90–94, Nov 2009.
3. G. Grün, *The development of the vertebrate retina: a comparative survey*. Springer Science & Business Media, 2012.
4. J. Dash and N. Bhoi, “A survey on blood vessel detection methodologies in retinal images,” in *IEEE International Conference on Computational Intelligence and Networks*, (Jabalpur, India), pp. 166–171, Jan 2015.
5. P. K. Saha, G. Borgefors, and G. S. di Baja, “A survey on skeletonization algorithms and their applications,” *Pattern Recognition Letters*, vol. 76, no. 1, pp. 3–12, 2016.
6. J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, “Ridge-based vessel segmentation in color images of the retina,” *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.
7. C. Lopez-Molina, G. V.-D. de Ulzurrun, J. Baetens, J. Van den Bulcke, and B. De Baets, “Unsupervised ridge detection using second order anisotropic Gaussian kernels,” *Signal Processing*, vol. 116, no. 1, pp. 55–67, 2015.
8. E. Smistad, “GPU-based airway tree segmentation and centerline extraction,” Master’s thesis, Institutt for Datateknikk Og Informasjonsvitenskap, 2012.
9. I. Sluimer, A. Schilham, M. Prokop, and B. van Ginneken, “Computer analysis of computed tomography scans of the lung: a survey,” *IEEE Transactions on Medical Imaging*, vol. 25, no. 4, pp. 385–405, 2006.
10. D. H. Chung and G. Sapiro, “Segmentation-free skeletonization of gray-scale images via PDEs,” in *IEEE International Conference on Image Processing*, (Quebec City, Canada), pp. 927–930, Sep 2000.
11. P. J. Yim, P. L. Choyke, and R. M. Summers, “Gray-scale skeletonization of small vessels in magnetic resonance angiography,” *IEEE Transactions on Medical Imaging*, vol. 19, no. 6, pp. 568–576, 2000.
12. T. Stosic and B. D. Stosic, “Multifractal analysis of human retinal vessels,” *IEEE Transactions on Medical Imaging*, vol. 25, no. 8, pp. 1101–1107, 2006.
13. A. M. Mendonca and A. Campilho, “Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 25, no. 9, pp. 1200–1213, 2006.
14. R. Annunziata, A. Kheirhah, P. Hamrah, and E. Trucco, “Scale and curvature invariant ridge detector for tortuous and fragmented structures,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Munich, Germany), pp. 588–595, Oct 2015.
15. S. R. Aylward, J. Jomier, S. Weeks, and E. Bullitt, “Registration and analysis of vascular images,” *International Journal of Computer Vision*, vol. 55, no. 2-3, pp. 123–138, 2003.

Table 7: Qualitative and quantitative results across samples of medical datasets (retinal vessels): DR HAGIS, HRF and non-biological datasets: CrackForest. The yellow box shows the Region Of Interest (ROI) in each image. The proposed, an average, and a popular combination have been shown in the last three columns respectively. The mean MHD for each dataset can be seen in the first three rows and the mean of all combinations' quality can be seen in the last row. For visibility, the output images have their intensity inverted.


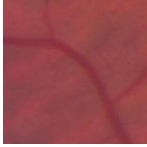




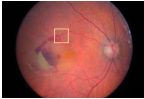
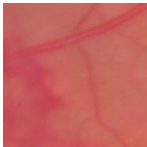


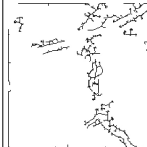



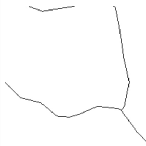
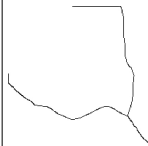
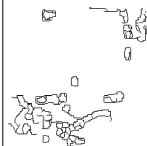
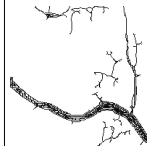

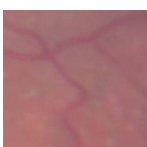




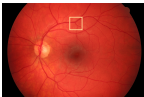

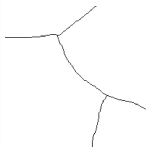
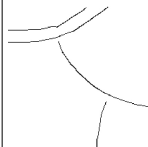

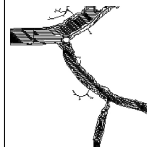
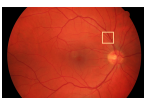
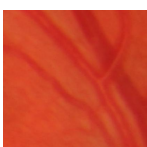


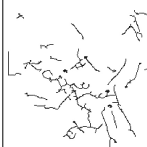


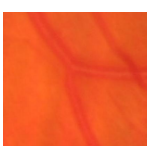


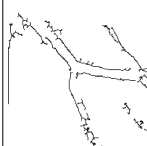
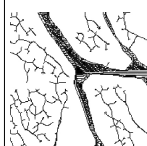



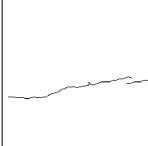
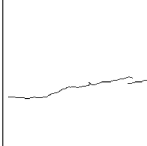
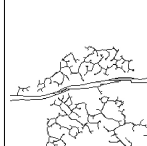

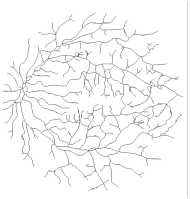

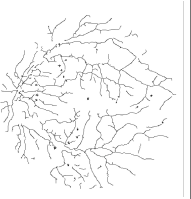
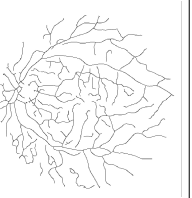

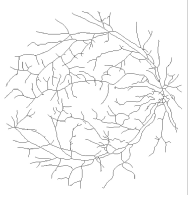
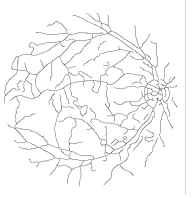
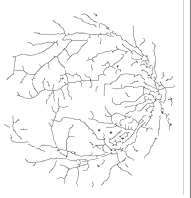
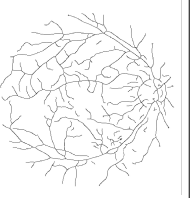
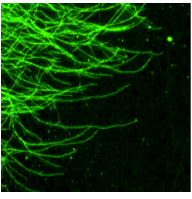


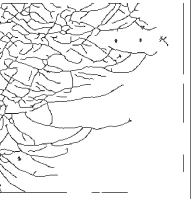
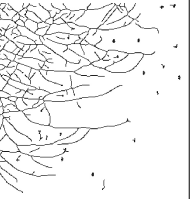
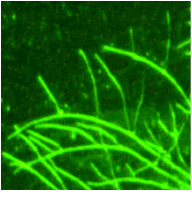



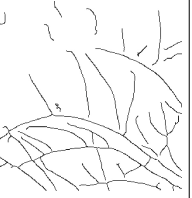

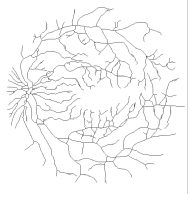
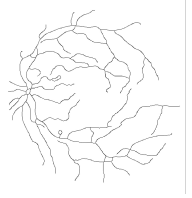
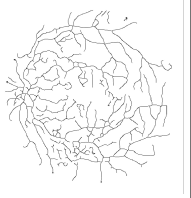

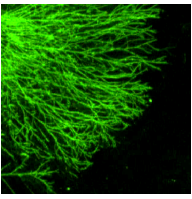

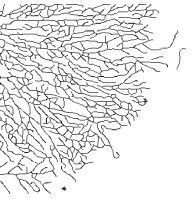
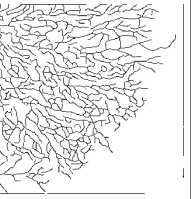
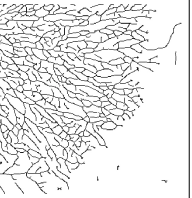
Dataset	Sample Image	ROI	Ground Truth	(Proposed) AGK & URDCS	(Average) PCT & URDCS	(Popular) RVR & URDCS
DR HAGIS						
						
						
						
	Mean MHD[ $\mu\text{m}$ ] for DR HAGIS			6.7307	16.008	11.6937
HRF						
						
						
	Mean MHD[ $\mu\text{m}$ ] for HRF			6.45	14.23	12.41
CrackForest						
	Mean MHD[ $\mu\text{m}$ ] for CrackForest			0.51	0.74	31.89
	Mean MHD[ $\mu\text{m}$ ] for All			4.58	10.32	18.06



Table 8: Qualitative and quantitative results across 20 images of the DRIVE dataset (samples of results are shown in the first two rows), 100 images of the GUF1-1 dataset (samples of results are shown in second two rows), and the last two rows show mixed data from both datasets, which reflect real-world data. The proposed, an average, and a popular combination have been shown in the last three rows respectively. For visibility, the output images have their intensity inverted and the GUF1-1 images are colorized with increased brightness and contrast (the original intensity can be seen in the crops in Figure 2b).

The mean of all combinations' MHD can be seen for each dataset.

	Images	Ground Truth	(Proposed) AGK & URDCS	(Average) PCT & URDCS	(Popular) RVR & URDCS
DRIVE					
					
	Mean MHD [pixel]		14.4798	17.6854	16.2639
GUF1-1					
					
	Mean MHD [pixel]		190.2182	1335.9433	628.6719
GUF1 & DRIVE					
					
	Mean MHD [pixel]		505.9168	1498.8783	907.89242

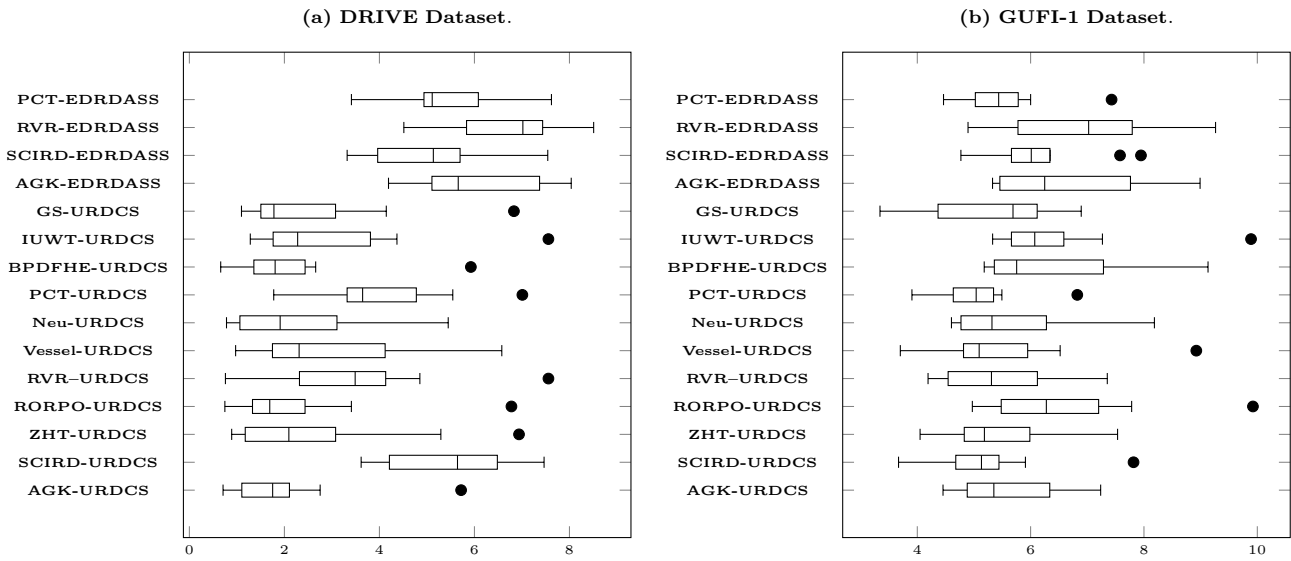


Fig. 3: Comparison between MHD scores of relevant selected combinations using cropped images from (a) the DRIVE dataset in the left box plot and (b) GUF1-1 dataset in the right box plot. These compared combinations are: **AGK-URDCS**, **SCIRD-URDCS**, **ZHT-URDCS**, **RORPO-URDCS**, **RVR-URDCS**, **Vessel-URDCS**, **Neu-URDCS**, **PCT-URDCS**, **BPDFHE-URDCS**, **IUWT-URDCS**, **GS-URDCS**, **AGK-EDRDASS**, **SCIRD-EDRDASS**, **RVR-EDRDASS** and **PCT-EDRDASS**.

16. Y. Zhou, A. Kaufman, and A. W. Toga, "Three-dimensional skeleton and centerline generation based on an approximate minimum distance field," *The Visual Computer*, vol. 14, no. 7, pp. 303–314, 1998.
17. E. Piuze, P. G. Kry, and K. Siddiqi, "Generalized helicoids for modeling hair geometry," *Computer Graphics Forum*, vol. 30, no. 2, pp. 247–256, 2011.
18. C. G. Willcocks, P. T. Jackson, C. J. Nelson, and B. Obara, "Extracting 3D parametric curves from 2D images of helical objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 9, pp. 1757–1769, 2017.
19. N. Strokina, T. Kurakina, T. Eerola, L. Lensu, and H. Kälviäinen, "Detection of curvilinear structures by tensor voting applied to fiber characterization," in *Scandinavian Conference on Image Analysis*, (Espoo, Finland), pp. 22–33, Jun 2013.
20. D. Maio and D. Maltoni, "Direct gray-scale minutiae detection in fingerprints," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 1, pp. 27–40, 1997.
21. A. M. López, F. Lumbreras, J. Serrat, and J. J. Villanueva, "Evaluation of methods for ridge and valley detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 4, pp. 327–335, 1999.
22. E. Bas and D. Erdogmus, "Principal curves as skeletons of tubular objects," *Neuroinformatics*, vol. 9, no. 2-3, pp. 181–191, 2011.
23. G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, and C. Pan, "Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3322–3337, 2017.
24. A. Sironi, V. Lepetit, and P. Fua, "Projection onto the manifold of elongated structures for accurate extraction," in *IEEE International Conference on Computer Vision*, (Santiago, Chile), pp. 316–324, Dec 2015.
25. W. Shen, K. Zhao, Y. Jiang, Y. Wang, Z. Zhang, and X. Bai, "Object skeleton extraction in natural images by fusing scale-associated deep side outputs," in *IEEE Conference on Computer Vision and Pattern Recognition*, (Las Vegas, US), pp. 222–230, Jun 2016.
26. Y. Ling, C. Yan, C. Liu, X. Wang, and H. Li, "Adaptive tone-preserved image detail enhancement," *The Visual Computer*, vol. 28, no. 6-8, pp. 733–742, 2012.
27. T. Lindeberg, "Edge detection and ridge detection with automatic scale selection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (San Francisco, US), pp. 465–470, Jun 1996.
28. A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Cambridge MA, USA), pp. 130–137, Oct 1998.
29. P. Perona, "Steerable-scalable kernels for edge detection and junction analysis," in *European Conference on Computer Vision*, (Santa Margherita Ligure, Italy), pp. 3–18, May 1992.
30. E. Meijering, M. Jacob, J.-C. Sarria, P. Steiner, H. Hirling, and M. Unser, "Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images," *Cytometry Part A*, vol. 58A, no. 2, pp. 167–176, 2004.
31. S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.
32. K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems*, (San Diego, California), pp. 474–485, 1994.
33. H. Zhu, F. H. Chan, and F. K. Lam, "Image contrast enhancement by constrained local histogram equalization," *Computer Vision and Image Understanding*, vol. 73, no. 2, pp. 281–290, 1999.



34. D. Sheet, H. Garud, A. Suveer, M. Mahadevappa, and J. Chatterjee, "Brightness preserving dynamic fuzzy histogram equalization," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, p. 10, 2010.
35. P. Joshi and S. Prakash, "Image enhancement with naturalness preservation," *The Visual Computer*, vol. 36, no. 1, pp. 71–83, 2020.
36. W. T. Freeman, E. H. Adelson, *et al.*, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991.
37. W. T. Freeman and E. H. Adelson, "Steerable filters for early vision, image analysis, and wavelet decomposition," in *IEEE International Conference on Computer Vision*, (Osaka, Japan), pp. 406–415, Dec 1990.
38. P.-L. Shui and W.-C. Zhang, "Noise-robust edge detector combining isotropic and anisotropic Gaussian kernels," *Pattern Recognition*, vol. 45, no. 2, pp. 806 – 820, 2012.
39. R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 4, pp. 532–550, 1987.
40. F. Zana and J.-C. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1010–1019, 2001.
41. O. Merveille, B. Naegel, H. Talbot, L. Najman, and N. Passat, "2D filtering of curvilinear structures by Ranking the Orientation Responses of Path Operators (RORPO)," *Image Processing On Line*, vol. 7, no. 1, pp. 246–261, 2017.
42. O. Merveille, H. Talbot, L. Najman, and N. Passat, "Curvilinear structure analysis by Ranking the Orientation Responses of Path Operators," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 2, pp. 304–317, 2018.
43. Ç. Sazak, C. J. Nelson, and B. Obara, "The multiscale bowler-hat transform for blood vessel enhancement in retinal images," *Pattern Recognition*, vol. 88, pp. 739–750, 2019.
44. Y. Sato, S. Nakajima, H. Atsumi, T. Koller, G. Gerig, S. Yoshida, and R. Kikinis, "3D multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," in *Computer Vision, Virtual Reality and Robotics in Medicine and Medical Robotics and Computer-Assisted Surgery Grenoble*, (Grenoble, France), pp. 213–222, Mar 1997.
45. T. Jerman, F. Pernuš, B. Likar, and Ž Špiclin, "Enhancement of vascular structures in 3D and 2D angiographic images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 9, pp. 2107–2118, 2016.
46. B. Obara, M. Fricker, D. Gavaghan, and V. Grau, "Contrast-independent curvilinear structure detection in biomedical images," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2572–2581, 2012.
47. P. Kovsi, "Phase congruency detects corners and edges," in *The Australian Pattern Recognition Society Conference*, (Brisbane), pp. 309–318, Feb 2003.
48. P. Bankhead, C. N. Scholfield, J. G. McGeown, and T. M. Curtis, "Fast retinal vessel detection and measurement using wavelets and edge location refinement," *PloS One*, vol. 7, no. 3, p. 32435, 2012.
49. N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
50. M. H. J. Vala and A. Baxi, "A review on Otsu image segmentation algorithm," *International Journal of Advanced Research in Computer Engineering & Technology*, vol. 2, no. 2, pp. pp–387, 2013.
51. M. S. Nixon and A. S. Aguado, *Feature extraction & image processing for computer vision*. Academic Press, 2012.
52. S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1522–1531, 2000.
53. X. Jiang and D. Mojon, "Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 1, pp. 131–137, 2003.
54. M. Sezgin *et al.*, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 146–168, 2004.
55. H. Blum and R. N. Nagel, "Shape description using weighted symmetric axis features," *Pattern Recognition*, vol. 10, no. 3, pp. 167–180, 1978.
56. N. D. Cornea, D. Silver, X. Yuan, and R. Balasubramanian, "Computing hierarchical curve-skeletons of 3d objects," *The Visual Computer*, vol. 21, no. 11, pp. 945–955, 2005.
57. L. Wade and R. E. Parent, "Automated generation of control skeletons for use in animation," *The Visual Computer*, vol. 18, no. 2, pp. 97–110, 2002.
58. M. S. Hassouna and A. A. Farag, "Robust centerline extraction framework using level sets," in *IEEE Conference on Computer Vision and Pattern Recognition*, (London, UK), pp. 458–465, Oct 2005.
59. M. S. Hassouna and A. A. Farag, "Variational curve skeletons using gradient vector flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2257–2274, 2009.
60. K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker, "Shock graphs and shape matching," *International Journal of Computer Vision*, vol. 35, no. 1, pp. 13–32, 1999.
61. K. Siddiqi, S. Bouix, A. Tannenbaum, and S. W. Zucker, "The Hamilton-Jacobi skeleton," in *IEEE International Conference on Computer Vision*, vol. 2, (Kerkyra, Greece), pp. 828–834, Sep 1999.
62. W. H. Hesselink and J. B. Roerdink, "Euclidean skeletons of digital image and volume data in linear time by the integer medial axis transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2204–2217, 2008.
63. A. Telea and J. J. van Wijk, "An augmented fast marching method for computing skeletons and centerlines," in *Proceedings of the Symposium on Data Visualisation*, (Barcelona, Spain), pp. 251–260, May 2002.
64. P. Maragos and R. Schafer, "Morphological skeleton representation and coding of binary images," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 5, pp. 1228–1244, 1986.
65. T. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Communications of the ACM*, vol. 27, no. 3, pp. 236–239, 1984.
66. Y.-S. Chen and W.-H. Hsu, "A modified fast parallel algorithm for thinning digital patterns," *Pattern Recognition Letters*, vol. 7, no. 2, pp. 99–106, 1988.
67. L. B. Boudaoud, A. Sider, and A. Tari, "A new thinning algorithm for binary images," in *International Conference on Control, Engineering & Information Technology*, (Tlemcen, Algeria), pp. 1–6, May 2015.
68. L. Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms,"

- IEEE Transactions on Image Processing*, vol. 2, no. 2, pp. 176–201, 1993.
69. C. Arcelli and G. S. Di Baja, “Finding local maxima in a pseudo-Euclidian distance transform,” *Computer Vision, Graphics, and Image Processing*, vol. 43, no. 3, pp. 361–367, 1988.
  70. G. Borgefors, “Centres of maximal discs in the 5-7-11 distance transform,” in *Scandinavian Conference on Image Analysis*, vol. 1, pp. 105–105, 1993.
  71. V. Chatzis and I. Pitas, “A generalized fuzzy mathematical morphology and its application in robust 2-D and 3-D object representation,” *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1798–1810, 2000.
  72. O. Sharma, D. Mioc, and F. Anton, “Voronoi diagram based automated skeleton extraction from colour scanned maps,” in *IEEE International Symposium on Voronoi Diagrams in Science and Engineering*, (Banff, Canada), pp. 186–195, Jul 2006.
  73. F. Corson, *Quelques aspects physiques du développement végétal*. PhD thesis, Université Pierre et Marie Curie-Paris VI, 2008.
  74. C. G. Willcocks and F. W. Li, “Feature-varying skeletonization,” *The Visual Computer*, vol. 28, no. 6-8, pp. 775–785, 2012.
  75. A. Sironi, E. Türetken, V. Lepetit, and P. Fua, “Multi-scale centerline detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 7, pp. 1327–1341, 2016.
  76. W. Shen, X. Bai, Z. Hu, and Z. Zhang, “Multiple instance subspace learning via partial random projection tree for local reflection symmetry in natural images,” *Pattern Recognition*, vol. 52, pp. 306–316, 2016.
  77. W. Shen, K. Zhao, Y. Jiang, Y. Wang, X. Bai, and A. Yuille, “Deepskeleton: Learning multi-task scale-associated deep side outputs for object skeleton extraction in natural images,” *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5298–5311, 2017.
  78. G. Wang, G. Van Stappen, and B. De Baets, “Automated artemia length measurement using u-shaped fully convolutional networks and second-order anisotropic gaussian kernels,” *Computers and Electronics in Agriculture*, vol. 168, p. 105102, 2020.
  79. T. Lindeberg, “Feature detection with automatic scale selection,” *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79–116, 1998.
  80. M. P. Do Carmo, *Differential Geometry of Curves and Surfaces: Revised and Updated Second Edition*. Courier Dover, 2016.
  81. C. Steger, “An unbiased detector of curvilinear structures,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 2, pp. 113–125, 1998.
  82. M. Jacob and M. Unser, “Design of steerable filters for feature detection using canny-like criteria,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1007–1019, 2004.
  83. S. S. Alharbi, C. G. Willcocks, P. T. Jackson, H. F. Alhasan, and B. Obara, “Sequential graph-based extraction of curvilinear structures,” *Signal, Image and Video Processing*, vol. 13, no. 5, pp. 941–949, 2019.
  84. M. Niemeijer, J. Staaf, B. van Ginneken, M. Loog, M. D. Abramoff, et al., “Comparative study of retinal vessel segmentation methods on a new publicly available database,” in *SPIE Medical Imaging*, vol. 5370, (San Diego, California), pp. 648–656, Feb 2004.
  85. P. Kovsi, “Image features from phase congruency,” *Videre: Journal of Computer Vision Research*, vol. 1, no. 3, pp. 1–26, 1999.
  86. S. Holm, G. Russell, V. Nourrit, and N. McLoughlin, “DR HAGIS - a fundus image database for the automatic extraction of retinal surface vessels from diabetic patients,” *Journal of Medical Imaging*, vol. 4, no. 1, p. 014503, 2017.
  87. A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, “Robust vessel segmentation in fundus images,” *International Journal of Biomedical Imaging*, vol. 2013, no. 11, pp. 1–11, 2013.
  88. Y. Shi, L. Cui, Z. Qi, F. Meng, and Z. Chen, “Automatic road crack detection using random structured forests,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3434–3445, 2016.
  89. D. E. Goldberg and J. H. Holland, “Genetic algorithms and machine learning,” *Machine Learning*, vol. 3, no. 2, pp. 95–99, 1988.
  90. A. R. Conn, N. I. Gould, and P. Toint, “A globally convergent augmented lagrangian algorithm for optimization with general constraints and simple bounds,” *SIAM Journal on Numerical Analysis*, vol. 28, no. 2, pp. 545–572, 1991.
  91. R. Klette and P. Zamperoni, “Measures of correspondence between binary patterns,” *Image and Vision Computing*, vol. 5, no. 4, pp. 287–295, 1987.
  92. A. Baddeley, “Errors in binary images and an lp version of the hausdorff metric,” *Nieuw Archief voor Wiskunde*, vol. 10, no. 4, pp. 157–183, 1992.
  93. D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, “Comparing images using the Hausdorff distance,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, 1993.
  94. Y. Lu, C. L. Tan, W. Huang, and L. Fan, “An approach to word image matching based on weighted Hausdorff distance,” in *International Conference on Document Analysis and Recognition*, (Seattle, US), pp. 921–925, Sep 2001.
  95. C. Zhao, W. Shi, and Y. Deng, “A new Hausdorff distance for image matching,” *Pattern Recognition Letters*, vol. 26, no. 5, pp. 581–586, 2005.
  96. C. F. Olson and D. P. Huttenlocher, “Automatic target recognition by matching oriented edge pixels,” *IEEE Transactions on Image Processing*, vol. 6, no. 1, pp. 103–113, 1997.
  97. D. M. Mount, N. S. Netanyahu, and J. Le Moigne, “Efficient algorithms for robust feature matching,” *Pattern Recognition*, vol. 32, no. 1, pp. 17–38, 1999.
  98. O.-K. Kwon, D.-G. Sim, and R.-H. Park, “Robust Hausdorff distance matching algorithms using pyramidal structures,” *Pattern Recognition*, vol. 34, no. 10, pp. 2005–2013, 2001.
  99. M.-P. Dubuisson and A. K. Jain, “A modified Hausdorff distance for object matching,” in *International Conference on Pattern Recognition, Computer Vision and Image Processing*, (Jerusalem), pp. 566–568, Oct 1994.
  100. B. Takacs, “Comparing face images using the modified Hausdorff distance,” *Pattern Recognition*, vol. 31, no. 12, pp. 1873–1881, 1998.
  101. K.-H. Lin, B. Guo, K.-M. Lam, and W.-C. Siu, “Human face recognition using a spatially weighted modified Hausdorff distance,” in *IEEE International Symposium on Intelligent Multimedia, Video and Speech Processing*, pp. 477–480, 2001.
  102. C.-B. Yu, H.-F. Qin, Y.-Z. Cui, and X.-Q. Hu, “Finger-vein image recognition combining modified Hausdorff distance with minutiae feature matching,” *Interdisciplinary Sciences: Computational Life Sciences*, vol. 1, no. 4, pp. 280–289, 2009.

- 
103. P. P. Sarangi, M. Panda, B. P. Mishra, and S. Dehuri, "An automated ear localization technique based on modified Hausdorff distance," in *International Conference on Computer Vision and Image Processing*, (Hong Kong), pp. 229–240, May 2017.
  104. P. E. Shrout and J. L. Fleiss, "Intraclass correlations: uses in assessing rater reliability," *Psychological Bulletin*, vol. 86, no. 2, p. 420, 1979.
  105. R. B. McCall and J. Kagan, "Fundamental statistics for psychology," tech. rep., Harcourt Brace Jovanovich New York, 1975.
  106. Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 1, no. 1, pp. 289–300, 1995.

Table 9: Comparing the mean of the MHD between the GT and the extracted ridges, for different enhancement and ridge detector combinations using images from DRIVE. Lighter coloured cells indicate smaller mean of MHD scores, which is similar to their ground truth, whereas darker blue indicates a higher mean of MHD scores. The last four rows and columns show the GM, ICC1, LRT and BH(p-value) across all enhancement and ridge detection methods respectively.

Enh/Seg		GS	CLAHE	Vessel	LTV	PCTV	CLHE	ZTH	AT	PCTN	BPDFHE	HT	PCT	IUWT	Neu	AGK	SCRD	RVR	RORFO	GM	ICC1	p-value	BH(p-value)
Skeleton	MT	5.70	5.19	4.38	5.18	5.07	6.18	3.33	4.36	5.38	17.02	18.30	5.38	5.38	3.65	3.65	4.39	3.41	5.18	7.07	0.36	$4.33 \cdot 10^{-2}$	$3.46 \cdot 10^{-2}$
	ZST	12.49	6.31	4.38	5.18	5.07	6.18	3.33	4.36	5.38	17.02	18.30	5.38	5.38	3.65	3.65	4.39	3.41	5.18	7.07	0.36	$4.33 \cdot 10^{-2}$	$3.46 \cdot 10^{-2}$
	MS	7.52	4.41	4.16	5.45	4.88	5.90	3.43	4.62	4.97	4.90	4.61	5.00	3.53	3.13	2.75	4.90	3.21	6.54	4.02	0.18	0.14	0.16
	EDRDASS	12.49	6.65	7.01	5.55	5.29	6.13	6.81	5.23	5.76	7.60	6.72	5.38	7.49	7.20	6.02	5.11	6.61	4.29	5.70	0.69	0.19	0.19
	UAD/CS	8.22	6.55	5.02	4.94	5.08	5.96	2.70	5.28	5.41	2.69	5.05	6.06	2.81	2.57	1.98	5.90	6.90	6.13	3.23	0.50	0.13	0.16
	SG	6.88	6.03	7.06	5.35	5.06	6.13	6.02	6.05	6.15	7.56	7.81	6.88	6.38	7.11	7.16	7.13	6.44	6.37	5.86	0.32	0.44	0.44
	VD	24.69	9.19	9.19	7.43	4.78	9.19	6.88	4.94	5.80	9.19	5.26	9.19	9.19	6.94	6.69	5.63	5.80	6.70	6.87	0.24	$2.21 \cdot 10^{-2}$	$4.41 \cdot 10^{-2}$
	AFMM	79.94	4.29	5.17	6.96	3.99	5.99	3.91	6.14	4.55	5.22	5.16	4.61	4.00	4.58	3.16	5.21	4.01	19.12	8.18	0.62	$2.15 \cdot 10^{-2}$	$4.41 \cdot 10^{-2}$
	GM	17.88	4.01	4.21	4.33	3.65	4.83	3.11	3.56	3.19	5.12	4.95	4.25	3.75	3.42	3.16	4.35	3.22	5.33	7.80	0.18	$6.67 \cdot 10^{-3}$	$2.26 \cdot 10^{-6}$
	p-value	$3.4 \cdot 10^{-3}$	0.19	0.17	0.44	$1.17 \cdot 10^{-2}$	$3.20 \cdot 10^{-9}$	$1.84 \cdot 10^{-2}$	0.14	$9.67 \cdot 10^{-2}$	$4.94 \cdot 10^{-3}$	$7.47 \cdot 10^{-3}$	0.26	$3.37 \cdot 10^{-9}$	0.18	$8.83 \cdot 10^{-2}$	$6.08 \cdot 10^{-4}$	$3.61 \cdot 10^{-4}$	0.13	0.19	0.19	0.19	0.19
BH(p-value)		$1.08 \cdot 10^{-2}$	0.21	0.21	0.44	$2.77 \cdot 10^{-2}$	$3.41 \cdot 10^{-8}$	$3.89 \cdot 10^{-2}$	0.19	0.15	$1.34 \cdot 10^{-2}$	$2.81 \cdot 10^{-3}$	0.27	$3.41 \cdot 10^{-8}$	0.21	0.15	$2.84 \cdot 10^{-3}$	$2.29 \cdot 10^{-3}$	0.13	0.19	0.19	0.19	0.19

Table 10: Comparing the mean of the MHD between the GT and the extracted ridges, for different enhancement and ridge detector combinations using images from GUF1-1. Darker coloured cells indicate a higher mean of MHD scores, which is similar to their ground truth, whereas lighter blue indicates smaller mean of MHD scores. The last four rows and columns show the GM, ICC1, LRT and BH(p-value) across all enhancement and ridge detection methods respectively.

Enh/Seg		GS	CLAHE	Vessel	LTV	PCTV	CLHE	ZTH	AT	PCTN	BPDFHE	HT	PCT	IUWT	Neu	AGK	SCRD	RVR	RORFO	GM	ICC1	p-value	BH(p-value)
Skeleton	MT	17.88	6.30	6.30	5.39	6.30	6.30	6.31	6.30	6.30	6.30	7.29	5.78	6.46	6.07	5.97	5.61	6.39	8.81	6.07	$5.33 \cdot 10^{-2}$	$4.31 \cdot 10^{-4}$	$1.54 \cdot 10^{-2}$
	ZST	20.50	6.31	6.30	7.08	6.77	6.76	6.66	7.78	6.31	7.78	7.64	6.02	6.96	6.84	6.52	6.03	7.61	8.81	6.71	0.43	$7.3 \cdot 10^{-3}$	$1.05 \cdot 10^{-2}$
	MS	8.99	5.86	6.01	6.16	6.15	6.23	5.99	6.91	5.65	6.92	6.82	5.57	6.39	6.07	5.73	5.93	6.47	6.19	5.49	0.23	$1.67 \cdot 10^{-2}$	$2.67 \cdot 10^{-2}$
	EDRDASS	8.22	6.95	6.51	6.37	6.06	6.94	6.59	6.30	6.02	7.81	7.08	6.02	7.72	7.43	6.84	6.03	6.90	6.63	5.96	0.20	$1.2 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$
	UAD/CS	6.88	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	5.86	0.32	0.44	0.44
	SG	6.88	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	6.03	5.86	0.32	0.44	0.44
	VD	20.23	9.19	9.19	8.70	5.89	9.19	6.41	6.79	6.19	9.19	6.73	9.19	9.11	6.67	6.79	6.06	6.48	11.51	7.27	0.50	$4.57 \cdot 10^{-2}$	$6.1 \cdot 10^{-2}$
	AFMM	40.78	5.55	6.78	5.37	5.05	5.46	6.08	6.81	5.80	7.07	7.04	5.33	5.64	6.33	5.71	5.35	6.68	28.82	7.80	$4.63 \cdot 10^{-28}$	$2.83 \cdot 10^{-7}$	$2.26 \cdot 10^{-6}$
	GM	11.02	6.78	6.12	4.78	4.43	5.15	4.65	4.94	4.77	5.43	4.35	4.01	4.38	4.36	4.51	4.33	4.75	3.81	7.80	0.18	$6.67 \cdot 10^{-3}$	$2.26 \cdot 10^{-6}$
	p-value	$3.9 \cdot 10^{-6}$	$4.8 \cdot 10^{-2}$	0.78	0.31	$7.5 \cdot 10^{-2}$	$8.39 \cdot 10^{-4}$	0.63	0.3	$2.64 \cdot 10^{-3}$	$1.16 \cdot 10^{-4}$	$1.38 \cdot 10^{-5}$	0.24	$4.99 \cdot 10^{-3}$	0.16	$5.49 \cdot 10^{-2}$	$1.54 \cdot 10^{-4}$	$2.15 \cdot 10^{-4}$	$5.42 \cdot 10^{-2}$	0.19	0.19	0.19	0.19
BH(p-value)		$7.4 \cdot 10^{-5}$	$8.69 \cdot 10^{-2}$	0.78	0.35	0.11	$2.28 \cdot 10^{-3}$	0.66	0.35	$6.27 \cdot 10^{-3}$	$5.5 \cdot 10^{-4}$	$1.31 \cdot 10^{-4}$	0.5	$3.16 \cdot 10^{-4}$	0.22	$8.69 \cdot 10^{-2}$	$5.85 \cdot 10^{-4}$	$6.8 \cdot 10^{-4}$	$8.69 \cdot 10^{-2}$	0.19	0.19	0.19	0.19

Table 11: Comparing the standard deviation of the MHD between the ground truth and the extracted ridges, for different enhancement and ridge detector combinations using images from DRIVE. Lighter coloured cells indicate a smaller standard deviation of MHD scores, which indicates robustness across different curve patterns, whereas darker blue indicates a higher standard deviation of MHD scores. The last row and column show the Root Mean Square (RMS) across all enhancement and ridge detection methods respectively.

Enh/Seg		OT	GS	CLAHE	Vessel	LTV	PCTV	CLHE	ZTH	AT	PCTN	BPDFHE	HT	PCT	IUWT	Neu	AGK	SCIRD	RVR	ROBPO	RMS
Skeleton	MT	69.36	2.39	2.10	1.91	1.70	1.22	1.11	2.21	2.56	1.53	26.71	26.97	1.66	1.80	1.68	1.75	1.07	2.50	1.95	18.22
	ZST	13.45	2.21	2.55	2.92	2.64	2.76	1.60	2.35	2.52	1.25	2.28	2.31	2.28	2.28	1.75	1.94	1.63	2.37	1.90	3.78
	MS	3.58	1.83	2.51	2.12	1.71	1.66	0.90	2.50	2.69	1.59	1.98	1.96	1.37	2.12	2.00	1.91	1.43	2.53	1.17	2.07
	FEDRASS	1.75	1.42	1.56	1.43	1.35	1.35	1.13	1.45	1.46	1.35	1.51	1.51	1.34	1.46	1.46	1.46	1.38	1.46	1.38	1.38
	URDGS	3.69	1.82	1.53	2.00	1.35	1.20	2.02	2.05	1.70	1.02	1.70	1.41	1.48	1.05	1.40	1.46	1.38	2.01	1.29	1.76
	SG	1.26	1.28	1.76	1.61	0.93	1.54	1.04	1.11	0.99	0.86	1.96	1.62	1.34	1.09	1.42	1.25	1.26	1.69	1.45	1.37
	VD	28.73	0.70	0.70	1.79	2.05	1.67	1.70	1.60	2.31	1.34	0.70	2.30	0.70	0.70	1.89	1.83	1.13	2.09	2.11	6.77
	AFMM	96.93	1.80	2.52	1.44	1.55	1.83	1.60	2.83	2.76	1.70	2.16	2.16	1.61	1.99	1.84	1.83	1.34	2.92	1.44	22.55
	RMS	46.73	1.92	2.12	2.1	1.82	1.84	1.45	2.21	2.4	1.5	10.26	10.44	1.63	1.94	1.8	1.8	1.43	2.39	5.58	11.43

Table 12: Comparing the standard deviation of the MHD between the ground truth and the extracted ridges, for different enhancement and ridge detector combinations using images from GUPF-1. Lighter coloured cells indicate a smaller standard deviation of MHD scores, which indicates robustness across different curve patterns, whereas darker blue indicates a higher standard deviation of MHD scores. The last row and column show the Root Mean Square (RMS) across all enhancement and ridge detection methods respectively.

Enh/Seg		OT	GS	CLAHE	Vessel	LTV	PCTV	CLHE	ZTH	AT	PCTN	BPDFHE	HT	PCT	IUWT	Neu	AGK	SCIRD	RVR	ROBPO	RMS
Skeleton	MT	8.78	1.58	1.54	1.41	1.12	1.19	1.72	1.39	1.08	1.49	1.89	1.93	1.23	1.40	1.40	0.90	0.96	0.80	2.25	2.46
	ZST	10.77	1.80	2.12	1.76	2.47	2.15	2.21	1.53	1.73	1.48	2.56	2.59	1.75	2.00	1.87	1.90	1.34	2.63	8.06	3.63
	MS	1.46	1.37	1.60	1.41	0.85	1.20	1.19	1.35	1.05	1.35	1.36	1.72	0.93	1.69	1.47	1.22	1.08	1.40	1.31	1.34
	FEDRASS	1.46	1.37	1.60	1.41	0.85	1.20	1.19	1.35	1.05	1.35	1.36	1.72	0.93	1.69	1.47	1.22	1.08	1.40	1.31	1.34
	URDGS	1.77	1.24	1.29	1.42	1.35	0.97	1.31	0.98	1.24	1.37	1.32	1.66	0.88	1.34	1.17	0.89	1.12	1.02	1.35	1.27
	SG	1.81	1.22	2.01	1.09	1.43	1.20	1.43	1.78	1.10	1.20	1.98	1.33	1.74	2.08	2.02	2.02	1.39	1.18	1.52	1.59
	VD	12.73	0.70	0.70	1.08	0.99	0.93	0.70	1.35	1.17	1.31	0.70	1.44	0.70	0.86	1.48	1.50	1.07	1.46	5.41	3.34
	AFMM	34.85	1.22	1.42	1.18	1.44	0.94	0.73	1.24	1.31	1.33	1.30	1.33	1.08	0.69	1.27	1.42	0.75	1.29	18.93	9.17
	RMS	15.02	1.44	1.65	1.45	1.53	1.37	1.52	1.47	1.39	1.45	1.79	1.86	1.3	1.58	1.62	1.52	1.17	1.58	8.15	4.18